

---

**S&P Global**

Market Intelligence

# Kensho Link Developer's Guide

**On-Demand Access**



## Client Support

Email	<a href="mailto:support.api.mi@spglobal.com">support.api.mi@spglobal.com</a>
Client Support Website	<a href="http://support.marketplace.spglobal.com">support.marketplace.spglobal.com</a>

# Contents

- 1 Document Overview..... 1**
  - 1.1 Introduction ..... 1
  - 1.2 Document Purpose ..... 1
  - 1.3 Client Support ..... 1
  
- 2 EDX FTP Access ..... 2**
  - 2.1 Firewall Rules..... 2
  - 2.2 Login Configurations ..... 2
  - 2.3 EDX URL and IP Addresses ..... 3
  - 2.4 Secure FTP Software..... 4
  - 2.5 Login Credentials ..... 4
  - 2.6 Directories ..... 5
  
- 3 Input File Formats ..... 6**
  - 3.1 File Name ..... 6
  - 3.2 File Delimiters ..... 6
  - 3.3 Input File Components..... 7
  - 3.4 Special/Foreign Character Handling ..... 7
  
- 4 Output File Formats ..... 7**
  - 4.1 Output File Name ..... 7
  - 4.2 Output File Components ..... 8
  
- 5 Examples of Input Files and Resulting Output Files ..... 9**
  
- 6 Common Error Responses..... 12**
  - 6.1 File Submissions ..... 12
  - 6.2 Individual Records..... 13
  - 6.3 Full List of Error Responses..... 13
  
- 7 Standard Account Attributes ..... 13**
  
- 8 Kensho Link Model ..... 14**
  - 7.1 Model..... 14
  - 7.2 Confidence Score..... 14
  
- 9 Revision History ..... 15**

# 1 Document Overview

## 1.1 Introduction

Kensho Link over API Drive is an enterprise on-demand platform that you can use to submit your unique portfolio of company names. Your custom portfolio data is returned to you on our FTP server as a delimited flat file, which can be imported into a data management system such as Microsoft Access, Microsoft Excel, SQL, and ORACLE.

## 1.2 Document Purpose

This document provides information for clients who subscribe to Kensho Link over API Drive. This guide provides you with:

- an understanding of the S&P Global Enterprise Data Exchange (EDX) platform using the TCP/IP protocol FTP
- the requirements and steps to upload and download files on the EDX platform
- formatting requirements for submitting files to S&P Global Market Intelligence
- the format in which you will receive files from S&P Global Market Intelligence
- a description of the Machine Learning model behind Kensho Link

## 1.3 Client Support

We are committed to quality products and customer service. Any time you have questions about your data, you will find our Client Support Center to be an important resource. For 24/7 support, email [support.api.mi@spglobal.com](mailto:support.api.mi@spglobal.com).

## 2 EDX FTP Access

Enterprise Data eXchange ("EDX") is S&P Global's enterprise level, scalable, reliable and highly available data delivery platform. EDX offers enhanced security for sensitive file transfers and provides support for all major secure file transfer standards including SFTP (FTP over SSH) or FTPS (FTP over TLS).

### 2.1 Firewall Rules

If your IT department needs to create new rules on your firewall to allow access to EDX, please make sure the new firewall rules include both Production and Disaster Recovery IP addresses.

### 2.2 Login Configurations

Your account team will work with you to determine the appropriate login configuration for SFTP or FTPS. Your account team will ask you for the.

- Remote Host or URL hosting SFTP or FTPS Services
- Remote Port
- Remote Username and Password or Remote Username and SSH Authorized User Key
- Remote Directory

## 2.3 EDX URL and IP Addresses

S&P Global data is delivered through a secured FTP site. You may access your data using SFTP or FTPS.

The domain for our EDX FTP site is **edx.spglobal.com**.

**Note:** We recommend that you use a domain name rather than an IP address. In the case of failure, we will automatically roll over to our back-up servers. As a result, if you are using an IP address, you will not have access to your files.

Following is a list of protocols, URLs, and IP addresses for the EDX FTP site.

EDX FTP URLs	Protocols	Port *	Production-1 IP Address	Production-2 IP Address
edx.spglobal.com	SFTP (SSH)	22	192.152.157.18	12.33.127.185
ftps-edx.spglobal.com	FTPS (TLS)	21 (Control Port)	192.152.157.19	12.33.127.186

\* The actual port numbers used do vary based on the Secure FTP selected. You can find them below:

### SFTP (Secure File Transfer Protocol):

- SSH Port: 22

### SFTP (Password Less/Public Key):

- SSH Port: 22

If you need to switch to **SFTP (Public Key)**, please contact [support.api.mi@spglobal.com](mailto:support.api.mi@spglobal.com) with the following information:

- Your firm's SSH Key.
- Request for S&P Global host key.
- Additional Details:
  - Contact Name
  - Telephone Number
  - Company Name
  - Business Unit/Department
  - FTP Account Usernames(s)
  - Names of other users at your firm accessing data via this account

### FTPS (FTP over TLS):

- TCP ports (For FTPS only):
  - 21 (Control Port)
  - 2001 to 2200 (Data Port) - All ports in this range need to be open

## 2.4 Secure FTP Software

In order to use Secure FTP, a compliant software package will be necessary. There are many available on the market from third parties (e.g. CuteFTP, FileZilla, OpenSSH suite, Putt/Psftp, Tectia SSH Client, WSFTP). EDX does not recommend or support third party software and thus, does not direct users to specific software packages.

## 2.5 Login Credentials

In order to access the EDX FTP server, a valid user name and password are required. You will receive an automated email with your username once your account is created. To obtain your password, please contact the API support team. Please keep in mind the following:

### User ID

- User IDs are case sensitive.
- User IDs can be customized, but there are some restrictions:
- Must be between 5 and 12 characters
- Cannot have spaces or symbols

### Password

- Passwords are case sensitive.
- Passwords can be customized, but there are some restrictions:
- Minimum of six characters.
- One character must be upper case, lower case and numeric.
- Underscores and symbols are allowed.
- Passwords are secure and cannot be retrieved by S&P Global Market Intelligence staff. If the password is lost, please contact Client Support to request a new password.

You will be automatically logged out after 10 minutes of inactivity.

## **2.6 Directories**

Your EDX site will display three folders:

### **2.6.1 Outbox**

Outbox – Kensho Link over API Drive Input Files

You may submit your Kensho Link Request Files to the Outbox folder any time the EDX Site is available.

### **2.6.2 Inbox**

Inbox – Kensho Link over API Drive Output Files

Inbox – Kensho Link over API Drive Output Files

Kensho Link Response Files are delivered to your Inbox folder as soon as they are available. Response Files will remain in your Inbox 14 days and will then be permanently deleted.

Kensho Link Response files will be available in the Inbox folder following these processing times:

- Requests of up to 100k rows take on average 1.5 to 2 hours
- Requests larger than 100k and smaller than 500k rows take on average 4 to 6 hours

Please note, Kensho Link works on a queue based system. Requests can take significantly longer to complete, if multiple files are submitted ahead of yours.



## 3 Input File Formats

### 3.1 File Name

Your Input file name should always follow the naming convention below:

**KL\_INPUTFILE\_<ID\_TYPE>\_<NUMBER\_OF\_MATCHES>\_<PORTFOLIO>.txt.**

Ex: **KL\_INPUTFILE\_CIQ\_3.txt**

**NOTE: Kindly Save the File using Notepad++.**

The input file must always begin with KL\_INPUTFILE and end with .txt. Your Input file name should always contain the .txt extension.

- **ID\_TYPE:** You can pass either CIQ or MI based on your requirements. Entering CIQ as a parameter will return the CIQ Company ID, while MI will link your portfolio to the MI Company ID.
- **NUMBER\_OF\_MATCHES:** This parameter allows you to choose how many matches you want for each record submitted. The value can range in between [1 to 5]
- **PORTFOLIO:** This parameter is an optional field that allows you to choose a name/id for the file. This parameter will help you track the file when multiple files are dropped into EDX.

Example: KL\_INPUTFILE\_CIQ\_3.txt

### 3.2 File Delimiters

Pipes should be used as delimiters between header keys and between fields. Use an escape sequence “\” backslash before the delimiter if the delimiter is part of the input text.

Example: 100|Greenville \ | KW

## 3.3 Input File Components

To produce meaningful data, your Input File should contain at a minimum two fields. These two fields are mandatory and the request will error out if they are not present for all records.

- **UID** : alphanumeric ID of the entity, cannot not be null, must be unique
- **NAME**: name of entity, cannot be null

### 3.3.1 Input Header Keys

Below is a list of all the headers, including optional fields that you can use when requesting a match. These fields can be submitted for each record in addition to the mandatory fields of **UID** and **NAME**. These optional header fields are all nullable. You may select any order of header keys.

HEADER NAME	DATATYPE	LENGTH	HEADER TYPE	COMMENTS
UID	VARCHAR	300	Mandatory	UIDs have to be unique
NAME	VARCHAR	600	Mandatory	
ALIASES	VARCHAR	800	Optional	Multiple values can be passed, separated by a comma (,)
YEAR_FOUNDED	INTEGER		Optional	
URL	VARCHAR	300	Optional	
PHONE_NUMBER	VARCHAR	15	Optional	
ADDRESS	VARCHAR	500	Optional	
CITY	VARCHAR	100	Optional	
ZIPCODE	VARCHAR	100	Optional	
STATE	VARCHAR	100	Optional	Full state name
COUNTRY	VARCHAR	100	Optional	Full country name

Following is an example of a header row in an Input File reflecting optional header keys.

```
UID|NAME|ALIASES|YEAR_FOUNDED|URL|PHONE_NUMBER|ADDRESS|CITY|ZIPCODE|STATE|COUNTRY
```

## 3.4 Special/Foreign Character Handling

Support for foreign alphabets and special characters in the Kensho Link Over API Drive is possible. However, the file should be saved using **UTF-8 Encoding**.

## 4 Output File Formats

### 4.1 Output File Name

A single output file will be generated for each input file submitted. Your Output file will be delivered in a compressed.zip file. Your Output filename will begin with your EDX User Id and will follow the same naming convention used in your Input file.

#### Input

```
KL_INPUTFILE_CIQ_1.txt
```

```
KL_INPUTFILE_MI_3_PORTFOLIO_1.txt
```

## Output

```
<EDX USER_ID>_KL_INPUTFILE_CIQ_1_<TIMESTAMP>.zip
```

```
<EDX USER_ID>_KL_INPUTFILE_MI_3_PORTFOLIO_1<TIMESTAMP>.zip
```

## 4.2 Output File Components

Each request response will be written to the file as:

```
UID|NAME|ENTITY_ID|ENTITY_CONFIDENCE_SCORE|
```

- **UID:** This is the ID from your input file, which will allow you to link the results back to your database
- **NAME:** This is the entity name from your input file
- **ENTITY\_ID:** This is either the MI or the CIQ Company ID, depending on what was entered as a parameter in the file
- **ENTITY\_CONFIDENCE\_SCORE:** This score represents the level of confidence behind the match. For more information on this field, please refer to [section 6.2 titled Confidence Score](#).

If the algorithm couldn't find a match in the database based on the information given, the **ENTITY\_ID** and the **ENTITY\_CONFIDENCE\_SCORE** will not be present in the output files.

You will instead see the message '**No Match Available**' at the record level. Please note, this is not an error.

## 5 Examples of Input Files and Resulting Output Files

Below are samples of Input Files and the resulting Output Files for various ID TYPE and NUMBER OF MATCHES.

### Input

```
UID|NAME|ALIASES|YEAR_FOUNDED|URL|PHONE_NUMBER|ADDRESS|CITY|ZIPCODE|S  
TATE|COUNTRY  
1|Google|google,Goog|1998|https://google.com|512-376-1876|1600  
Amphitheatre Parkway|Mountain View|California|94035|United States
```

### Output

```
UID|NAME|ENTITY_ID|ENTITY_CONFIDENCE_SCORE  
1|GOOGLE INC|267349|100
```

### Input

In this example, an extra field is passed in the request row. The extra field (year founded) will not be considered as there is no header for the field.

```
UID|NAME|COUNTRY
```

```
10|GOOGLE|UNITED STATES|1998  
21|APPLE|UNITED STATES|1976  
24337|Motoexpert|UNITED STATES|2000  
2388|Entreiva|UNITED STATES|1989
```

### Output

```
UID|NAME|ENTITY_ID|ENTITY_CONFIDENCE_SCORE
```

```
10|GOOGLE|312932093|97  
21|APPLE|24937|96  
2388|Entreiva|NO MATCH AVAILABLE  
24337|Motoexpert|NO MATCH AVAILABLE
```

## Kensho Link Developer's Guide

### Input

In this example, we are asking the file to return 3 matches for each record

```
UID|NAME
1|Ceprotec
2|Cosmobutler
11AC|GCX Corporation
3|Strateco Resources Inc.
4|Xtelesis Corporation
5|Fiton Technologies Corp.
6|Pennex Aluminum Company LLC
Aa123|CarTrawler
7|Nucap Industries Inc.
8|Gumbiner Savett Inc.
9|Srinivas
10|GOOGLE
21|APPLE
24337|Motoexpert
2388|Entreiva
17905|bessermitfahren.de Twomoons
25504|Comeqtec Corp.
```

### Output

The output data is ordered by ascending UID and descending confidence score.

```
UID|NAME|ENTITY_ID|ENTITY_CONFIDENCE_SCORE
1|Ceprotec|NO MATCH AVAILABLE
10|GOOGLE|312932093|97
10|GOOGLE|53117157|94
10|GOOGLE|29096|85
11AC|GCX Corporation|6701585|100
11AC|GCX Corporation|274264094|85
11AC|GCX Corporation|98162|73
17905|bessermitfahren.de Twomoons|NO MATCH AVAILABLE
2|Cosmobutler|NO MATCH AVAILABLE
21|APPLE|24937|96
21|APPLE|275892751|89
21|APPLE|4310283|60
2388|Entreiva|NO MATCH AVAILABLE
24337|Motoexpert|NO MATCH AVAILABLE
25504|Comeqtec Corp.|NO MATCH AVAILABLE
3|Strateco Resources Inc.|3182443|100
3|Strateco Resources Inc.|252135602|5
3|Strateco Resources Inc.|391396|4
4|Xtelesis Corporation|7991008|99
5|Fiton Technologies Corp.|3110213|100
5|Fiton Technologies Corp.|595377756|15
5|Fiton Technologies Corp.|23135763|14
6|Pennex Aluminum Company LLC|46740630|100
```

## Kensho Link Developer's Guide

```
6|Pennex Aluminum Company LLC|20407530|4
6|Pennex Aluminum Company LLC|25118465|2
7|Nucap Industries Inc.|6906363|100
7|Nucap Industries Inc.|646932115|15
7|Nucap Industries Inc.|224538172|4
8|Gumbiner Savett Inc.|6833281|100
8|Gumbiner Savett Inc.|116734485|1
8|Gumbiner Savett Inc.|243820852|0
9|Srinivas|628385945|95
9|Srinivas|62442668|41
9|Srinivas|9898876|23
Aa123|CarTrawler|24739812|99
```

## 6 Common Error Responses

### 6.1 File Submissions

It is possible to receive the following error responses when submitting an Input File:

Error	Explanation
Annual limit exceeded. The file has not been processed, please reach out to Client Support	When the annual limit of records is exceeded, an empty response file is provided. The error message will appear in the response file.
Not entitled. Please reach out to Client Support.	Please reach out to Client Support
ERROR PROCESSING REQUEST	The file hasn't been submitted, please reach out to Client Support.
Please specify identifier type in the file name	For a list of valid identifier types, please refer to <a href="#">section 3.1 File Name</a>
Please specify number of results in the file name	For a list of valid values, please refer to <a href="#">section 3.1 File Name</a>
Mandatory HEADERS- ID, Name missing	UID and NAME are mandatory headers and fields. Please refer to <a href="#">section 3.3.1 Header Keys</a> for a complete list of headers.
Duplicate HEADER [value1,value2]	Headers cannot be repeated.
Invalid HEADER [values]	For a list of valid headers, please refer to <a href="#">section 3.3.1 Input Header Keys</a>
Invalid DELIMITER	The file accepts pipes ( ) as delimiters
Invalid file format	More than one format error present in the file. Please refer to <a href="#">Input File Formats</a> for formatting standards
Input File empty	Input File empty
Maximum 1 million rows allowed	The input file can have a maximum of 500k records.

## 6.2 Individual Records

It is possible to receive the following error responses at a record level when submitting a request:

Error	Explanation
ROW – [value]. Please specify mandatory column-UID	This error is at a record level. The record mentioned in the error is missing a value for the mandatory column UID
ROW - [value]. UID not unique	This error is at a record level. The record mentioned in the error has a duplicate value for UID. UIDs have to be unique throughout the file
ROW - [value]. Please specify mandatory column-Name	This error is at a record level. The record mentioned in the error is missing a value for the mandatory column NAME

## 6.3 Full List of Error Responses

For a full list of responses, please refer to the Error Messages Guide under S&P Capital IQ API on the [Support Center](#).

## 7 Standard Account Attributes

Throttling Limits	
KENSHO_LINK_FILE_REQUEST_LIMIT	500000
KENSHO_LINK_ANNUAL_REQUEST_LIMIT	1000000



## 8 Kensho Link Model

### 7.1 Model

When you submit a file for linking, that file is sent to our model server. There, the linking process starts with cleaning and standardizing the data. Next, our machine learning model identifies a set of likely candidates for each record using quick similarity statistics before computing an overall match score for each candidate by computing and weighing similarities between every field in each record.

### 7.2 Confidence Score

While this match score doesn't correspond exactly to a probability, it does reflect the model's confidence.

Internally, we generally interpret confidence scores as follows:

- Confidence score is > 90%: this is a good link and may be added to an auto-linking queue
- Confidence score is < 10%: even the best link found was bad so there is likely no link for this entity
- No match found: no probable links for this entity could be found in the database. These records will be returned to you with a 'no match available' response.

Anything in-between may require human examination, as the model is unsure whether or not the link it identified is correct. This occurs when the model is able to find a similar but-just-dissimilar-enough record for the user provided record.

## 9 Revision History

The changes made to this document include the following:

Version	Date	Changes
1.0	04/28/2020	Initial version
1.1	04/29/2020	Updated language for section 2.6.2 Inbox Removed section 2.6.3 Archive Removed the error 'Not Applicable' in section 6.1 File Submissions
1.2	12/08/2020	Updated Input file detail and added saving recommendation to Notepad++ in section 3.1 Added Special/Foreign Character Handling in section 3.4 Added Standard Account Attributes in section 7
1.3	01/29/2021	Removed Notepad++ reference from section 3.4 for support for foreign alphabets and special characters.
1.4	03/12/2021	Updated API support email address and support website reference.
1.5	07/29/2021	Updated special signs and underscores on passwords are allowed.
1.6	05/19/2023	Removed delimiter at end of inputs and output file formats in sections 3 and 5

Copyright © 2023 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved. These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively, S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers, (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON "AS IS" BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence's opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence may provide index data. Direct investment in an index is not possible. Exposure to an asset class represented by an index is available through investable instruments based on that index. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global keeps certain activities of its divisions separate from each other in order to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain non- public information received in connection with each analytical process.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global's public ratings and analyses are made available on its Web sites, [www.standardandpoors.com](http://www.standardandpoors.com) (free of charge) and [www.ratingsdirect.com](http://www.ratingsdirect.com) (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at [www.standardandpoors.com/usratingsfees](http://www.standardandpoors.com/usratingsfees).